

Analytische Modellierung des Zeitverhaltens und der Verlustleistung von CMOS-Gattern

R. Geißler and H.-J. Pfeleiderer

Universität Ulm, Abteilung Allgemeine Elektrotechnik und Mikroelektronik, Albert-Einstein-Allee 43, D-89081 Ulm, Germany

Zusammenfassung. In modernen CMOS-Technologien werden die Verzögerungszeit, die Ausgangsflankensteilheit und der Querstrom eines Gatters sowohl durch die Lastkapazität als auch durch die Steilheit des Eingangssignals beeinflusst. Die heute verwendeten Technologiebibliotheken beinhalten Tabellenmodelle mit 25 oder mehr Stützpunkten dieser Abhängigkeiten, woraus durch Interpolation die benötigten Zwischenwerte berechnet werden. Bisherige Versuche, analytische Modelle abzuleiten beruhten darauf, den Querstrom zu vernachlässigen oder Transistorströme als stückweise linear anzunähern. Der hier gezeigte Ansatz beruht auf einer näherungsweise Lösung der Differentialgleichung, die aus den beiden Transistorströmen und einer Lastkapazität besteht und damit das Schaltverhalten eines Inverters beschreibt. Mit wenigen Technologieparametern können daraus für einen beliebig dimensionierten Inverter die für eine Timing- und Verlustleistungsanalyse notwendigen Größen berechnet werden. Das Modell erreicht bei einem Vergleich zu Referenzwerten aus SPICE Simulationen eine Genauigkeit von typischerweise 5%.

In modern CMOS-technologies the gate delay, output transition time and the short-circuit current depend on the capacitive load as well as on the input transition time. Today's technology libraries use table models with 25 or more samples for these dependencies. Intermediate values have to be calculated through interpolation. Attempts to derive analytical models are based on neglecting the short-circuit current or approximating it by piecewise linear functions. The approach shown in this paper provides an approximate solution for the differential equation describing the dynamic behavior of an inverter circuit. It includes the influence of both transistor currents and a single load capacitance. The required values for timing and power analysis can be calculated with a small set of technology parameters for an arbitrary designed inverter. Compared to reference values extracted from

SPICE simulations, the model achieves a typical precision of 5%.

1 Einleitung

Beim Entwurf von digitalen CMOS-Schaltungen ist es während und nach der Schaltungssynthese erforderlich, deren Zeitverhalten und ggf. deren Verlustleistung zu analysieren. Für ältere Technologien ($0,8\ \mu\text{m}$, $1,0\ \mu\text{m}$ und größer) wurden hierfür sehr einfache, lineare Modelle unter alleiniger Berücksichtigung der kapazitiven Last am Ausgang des jeweiligen Gatters herangezogen. In neueren und damit kleinere Strukturgrößen vorweisenden Technologien wird zusätzlich die Flankendauer des Eingangssignals berücksichtigt. Dazu werden üblicherweise die Elemente einer Technologiebibliothek entweder durch Analogsimulation oder durch Messung charakterisiert und diese Ergebnisse in Tabellenmodelle gespeichert.

Aufgrund der Tatsache, dass analytische Modelle zum besseren Verständnis der Einflüsse von Technologie- und Designparameter beitragen, wurden in der Vergangenheit verschiedene derartige Modelle entwickelt. Hedenstierna (1987) und Jeppson (1994) leiteten analytische Zusammenhänge mit heute nicht mehr zutreffenden Transistormodellen ab. Weitere Ansätze beruhten auf der Vernachlässigung des Inverter-Querstroms (Sakurai, 1990; Vemuru, 1994; Dutta, 1995), welche brauchbare Ergebnisse nur für kurze, d.h. schnelle Eingangsflanken lieferten. Mittels stückweise linearer Näherung des Inverter-Querstroms (Hirata, 1996; Bisdounis, 1998, 2000) und zusätzlicher Taylorreihen-Entwicklungen bzw. Linearisierungen konnten Lösungen für das Zeitverhalten und die Verlustleistung entwickelt werden. Andere Ansätze leiteten analytische Lösungen für den Fall einer schnellen Eingangsflanke ab und erweiterten diese durch Korrekturterme mit empirisch optimierten Parametern (Auvierge, 2000).

Der hier dargestellte Ansatz versucht eine analytische Näherungslösung aufzuzeigen, die mit wenigen und für eine

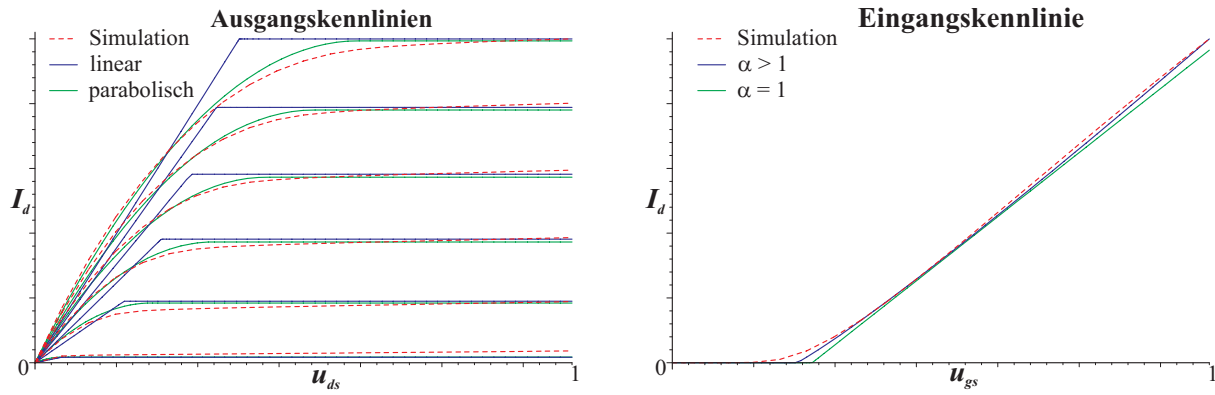


Abbildung 1. Varianten des alpha-Power Modells im Vergleich zu numerischen Kennlinien auf der Grundlage eines MM9 Modells.

Technologie konstanten Parametern die Analyse des Zeitverhaltens und der Verlustleistung ermöglicht.

2 Mathematische MOS-Modelle

Grundlage für die analytische Charakterisierung von digitalen CMOS-Schaltungen ist die mathematische Beschreibung des Großsignal-Verhaltens der Transistoren, d.h. des Drainstroms I_d als Funktion der beiden Spannungen U_{gs} und U_{ds} . Derartige Modelle können entweder durch Analyse der physikalischen Wechselwirkungen im Halbleiter (z.B. Sah, 1964) abgeleitet oder durch Parameteranpassung von einfachen mathematischen Funktionen (Sakurai, 1990) erstellt werden. Die Analyse physikalischer Wechselwirkungen erweist sich für heutige Technologien als äußerst komplex, weil hierbei verschiedene Kurzkanal-Effekte zu berücksichtigen sind. Die daraus resultierenden, umfangreichen Modellgleichungen können in numerisch arbeitenden Simulationswerkzeugen wie SPICE eingesetzt werden. Zur analytischen Bewertung von digitalen CMOS-Schaltungen sind diese Gleichungen jedoch ungeeignet.

Als Alternative zur physikalisch-analytischen Herleitung können Transistormodelle auch durch die Anpassung von einfachen, parametrisierten Funktionen (z.B. Ebenen, parabolische Flächen) an gemessene oder numerisch gewonnene Kennlinienfelder erstellt werden. Das alpha-Power Modell (Sakurai, 1990) ist ein derartiges Modell, das sich durch seine mathematisch einfache Beschreibung und einer für digitale Schaltungen brauchbaren Genauigkeit auszeichnet:

$$I_{D,n} = \begin{cases} k_{ln}(u_{gs} - n)^{\frac{\alpha}{2}} \cdot u_{ds} & \text{für } u_{ds} < k_{sn}/k_{ln}(u_{gs} - n)^{\frac{\alpha}{2}} \\ k_{sn}(u_{gs} - n)^{\alpha} & \text{für } u_{ds} \geq k_{sn}/k_{ln}(u_{gs} - n)^{\frac{\alpha}{2}} \end{cases} \quad (1)$$

Die Parameter k_{ln} , k_{sn} und α bezeichnen hierbei Design- und Technologiekonstanten, die u.a. von der Dimensionierung des Transistors abhängen. Die Größen u_{gs} , u_{ds} und n entsprechen den auf die Versorgungsspannung V_{dd} normierten Gate-Source-, Drain-Source- und Einsatzspannung. Abbildung 1 zeigt anhand der Eingangs- und Ausgangskennli-

nien einen Vergleich des alpha-Power Modells (blau dargestellt) mit einem numerisch, auf der Grundlage eines MM9 Modells (Velghe, 1993) simulierten Beschreibung (rot gestrichelt). Es zeigt sich eine gute Übereinstimmung für den Sättigungsbereich, wohingegen die Approximation im linearen Arbeitsbereich sehr grob ist.

Bedingt durch unterschiedliche Anforderungen an die Genauigkeit sind in der Literatur diverse Varianten des alpha-Power Modells entstanden:

- Verwendung einer parabolischen Funktion zur präziseren Modellierung im linearen Arbeitsbereich (Sakurai, 1991; Hirata, 1996) (Grüne Kurve in den Ausgangskennlinien von Abb. 1).
- Vereinfachung der Gleichungen durch die Definition $\alpha = 1$ (z.B. Auvergne, 2000). Aufgrund der Tendenz, dass der Exponent α für kleinere Strukturgrößen dem Wert 1 zustrebt (z.B. $\alpha = 1,1, \dots, 1,3$ für eine $0,35 \mu\text{m}$ Technologie), wird der damit eingeführte Fehler kontinuierlich kleiner und damit vernachlässigbar. Die resultierenden Modellgleichungen werden sowohl analytisch als auch numerisch günstiger. In Abb. 1 ist dieses Modell grün in den Eingangskennlinien dargestellt.
- Berücksichtigung eines mit U_{ds} ansteigenden Sättigungsstroms: Mittels eines multiplikativen Terms $I_D^* = I_D \cdot (1 + \lambda u_{ds})$ kann der Kanallängenmodulationsfaktor λ (Hirata, 1996) in das alpha-Power Modell eingebracht werden.

Die bisher aufgeführten Modellierungsansätze erfordern eine Unterscheidung des Arbeitspunkts in Sättigungs- und linearen Bereich. Um diese Unterscheidung und die damit implizierten Fallunterscheidungen bei der Analyse des Schaltverhaltens von CMOS-Schaltungen (vgl. Kap. 3) zu eliminieren, kann die folgende, für beide Arbeitsbereiche gemeinsam gültige Modellgleichung genutzt werden:

$$I_{D,n} = k_n \cdot (u_{gs} - n) \cdot \left(1 - a_n \cdot \exp\left(-u_{1,n} \frac{u_{ds}}{u_{gs}}\right) - b_n \cdot \exp\left(-u_{2,n} \frac{u_{ds}^2}{u_{gs}^2}\right) \right) \quad (2)$$

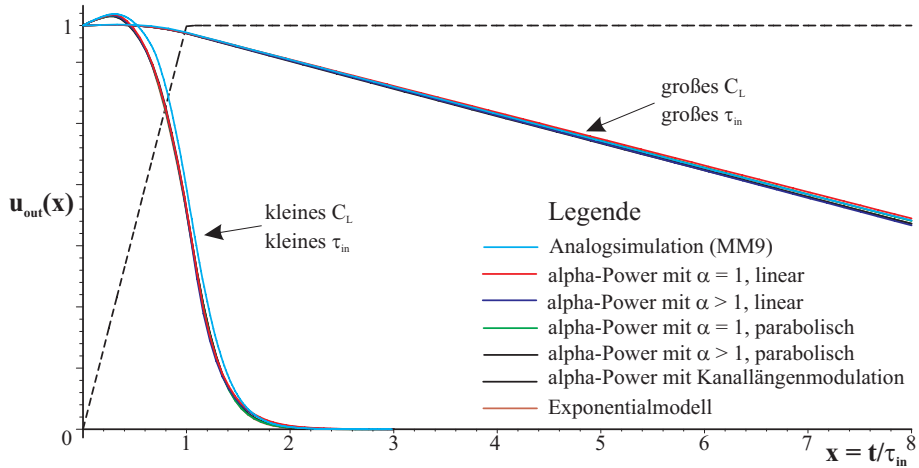


Abbildung 2. Numerisch berechnete Inverter-Ausgangssignale auf der Grundlage verschiedener mathematischer MOS-Modelle und Vergleich zur Analogsimulation.

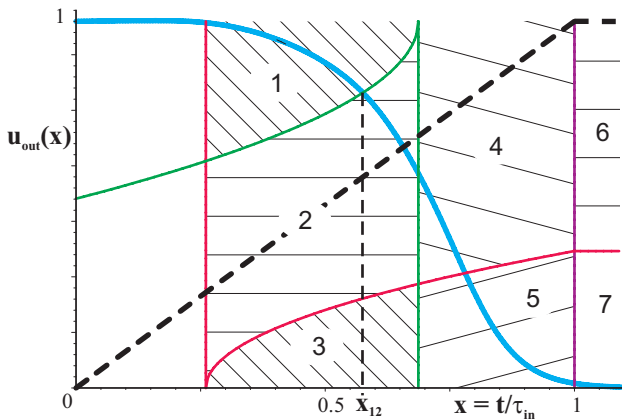


Abbildung 3. Untergliederung des Lösungsraums für das Inverter-Ausgangssignal.

Welches dieser Modelle eine hinreichende Genauigkeit bei vertretbarer Komplexität aufweist, zeigt die numerische Berechnung des Schaltverhaltens eines CMOS-Inverters und Vergleich zu einem als Referenz dienenden Signalverlauf aus einer Analogsimulation mit ELDO. Als Referenz hierzu diene ein MM9 Transistormodell mit Parametern einer 0,35 μm CMOS-Technologie. Abbildung 2 zeigt den Vergleich der berechneten Signalverläufe, wobei zur besseren Darstellbarkeit die Zeitachse auf die Flankendauer τ_{in} des Eingangssignals normiert wurde.

In Abb. 2 liegen die numerisch berechneten Inverter-Ausgangssignale mit den einfachen, mathematischen MOS-Modellen (alpha-Power und Exponential-Modell) annähernd deckungsgleich übereinander. Daher wurde für die folgende, analytische Untersuchung das mathematisch am einfachsten beschriebene MOS-Modell (alpha-Power Modell mit $\alpha = 1$ und linearer Beschreibung im linearen Arbeitsbereich) herangezogen.

3 Invertermodell

Für die analytische Betrachtung eines Schaltvorgangs an einem kapazitiv belasteten Inverter können unterschiedlich präzise Ersatzschaltbilder verwendet werden. Diese Ersatzschaltbilder unterscheiden sich in folgenden Aspekten:

- Behandlung des Kurzschlussstroms: Während die Vernachlässigung des Transistors, welcher den Kurzschlussstrom liefert, für schnelle Eingangsfanken sinnvoll erscheint, kann hierdurch bei langsamen Eingangsfanken und kleinen Lastkapazitäten ein nicht vernachlässigbarer Fehler entstehen.
- Parasitäre Transistor-Kapazitäten: Die Gate-Source- und Drain-Bulk-Kapazitäten werden i.d.R. als konstante Kapazitäten zwischen Eingang und Masse und zwischen Ausgang und Masse berücksichtigt. Die Unterschiede liegen in der Modellierung der Gate-Drain-Kapazitäten, entweder als additiver Anteil der oben genannten Ersatzkapazitäten oder als Koppelkapazität zwischen Ein- und Ausgang des Inverters.

Für die folgenden Betrachtungen soll ein Invertermodell herangezogen werden, das die beiden Drain-Ströme von NMOS- und PMOS-Transistor berücksichtigt, sowie die Gate-Drain-Kapazitäten auf reine Eingangs- und Ausgangs-Kapazitäten abbildet. Der Verlauf der Ausgangsspannung wird somit bei einer rampenförmigen Eingangsspannung durch folgende Differentialgleichung beschreiben:

$$I_{D,n} + I_{D,p} + \frac{C_L V_{dd}}{\tau_{in}} \frac{du_{out}}{dx} = 0 \quad \text{mit } x = \frac{t}{\tau_{in}} \quad (3)$$

Als Modell für die beiden Transistorströme $I_{D,n}$ und $I_{D,p}$ soll das alpha-Power Modell mit $\alpha = 1$ und linearer Beschreibung im linearen Arbeitsbereich eingesetzt werden.

Generell macht die abschnittsweise Definition des Drain-Stroms in Abhängigkeit von den Gate-Source- und Drain-Source-Spannungen eine abschnittsweise Lösung der Differentialgleichung (3) erforderlich. Dazu können, wie in Abb. 3

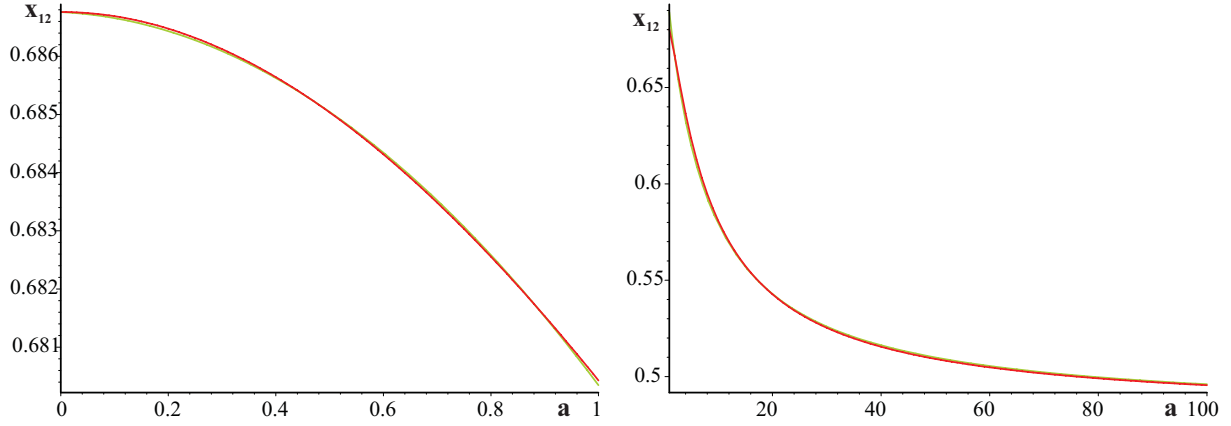


Abbildung 4. Numerische Lösung für x_{12} und Approximation.

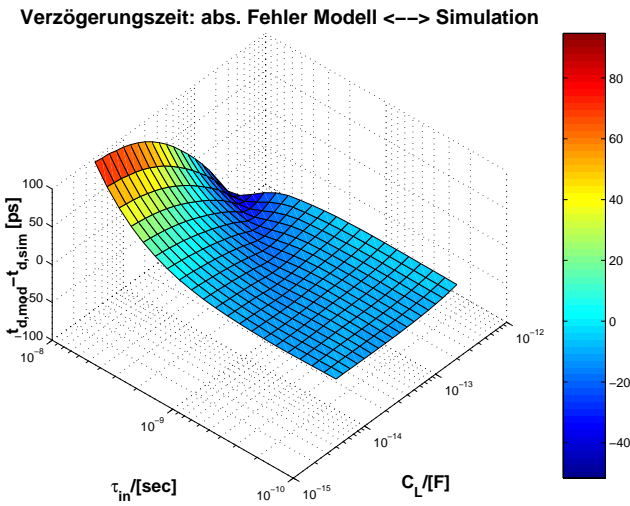


Abbildung 5. Absoluter Fehler zwischen Modell und Analogsimulation.

dargestellt, sieben Regionen abgegrenzt werden. Für die Analyse des Zeitverhaltens und des Kurzschlussstroms sind Lösungen für die Regionen 1, 2, 4 und 6 notwendig.

Essentiell wichtig ist eine hinreichend genaue Lösung für die Region 1 bzw. für den Schnittpunkt des Ausgangssignals mit der Grenzlinie zur Region 2. Allerdings existiert für die Differentialgleichung (3) in der Region 1 keine analytische Lösung. Das Problem liegt hierbei in der Lösung des folgenden Integrals:

$$Int = \int_{u=n}^x (u-n) \exp\left(-\frac{2}{3} \frac{\tau_{in} k_{lp}}{C_L V_{dd}} (1-u-p)^{\frac{3}{2}}\right) du \quad (4)$$

Abhilfe kann hier eine Taylorreihenentwicklung des Terms $(1-u-p)^{\frac{3}{2}}$ um den Punkt $u=0,5$ bis zum quadratischen Glied schaffen. Mit der daraus resultierenden Näherungslösung und der Beschreibung der Grenzkurve zwischen Region 1 und 2 kann nun der Schnittpunkt x_{12} des Ausgangssignals mit dieser Grenzkurve ermittelt werden. Dies kann

durch eine numerische Lösung der Gleichung (5) erfolgen.

$$\begin{aligned} & k_{sp} \sqrt{1-p-x} + \frac{1}{2} k_{sn} e^{\left(\frac{2}{3} \frac{\tau_{in} k_{lp} (1-p-x)^{\frac{3}{2}}}{C_L V_{dd}}\right)} \\ & \left(2\sqrt{2-4p} \left(e^{\left(\frac{1}{48} \frac{\sqrt{2-4p} k_{lp} \tau_{in} (23-56p+32p^2-36x+48xp+12x^2)}{(-1+2p)C_L V_{dd}}\right)}\right.\right. \\ & \left.\left.- e^{\left(\frac{1}{48} \frac{\sqrt{2-4p} k_{lp} \tau_{in} (12n^2-36n+48np+23-56p+32p^2)}{(-1+2p)C_L V_{dd}}\right)}\right)\right) \\ & + (3-4p-2n) \sqrt{\pi} e^{\left(-\frac{1}{12} \frac{\sqrt{2-4p} k_{lp} \tau_{in} (-1+2p)}{C_L V_{dd}}\right)} \left(\frac{1}{2}-p\right)^{\left(\frac{1}{4}\right)} \\ & \sqrt{\frac{k_{lp} \tau_{in}}{C_L V_{dd}}} \left(\operatorname{erf}\left(\frac{1}{4} \frac{(4p+2n-3)\sqrt{2}\sqrt{k_{lp} \tau_{in}}}{\sqrt{\sqrt{2-4p} C_L V_{dd}}}\right)\right. \\ & \left.- \operatorname{erf}\left(\frac{1}{4} \frac{(4p+2x-3)\sqrt{2}\sqrt{k_{lp} \tau_{in}}}{\sqrt{\sqrt{2-4p} C_L V_{dd}}}\right)\right) \Big) = 0 \quad (5) \end{aligned}$$

Nachdem eine derart rechenintensive Lösung für die Analyse von umfangreichen Digitalerschaltungen sich als ungünstig erweist, ist die Herleitung einer weiteren Näherung zur Vereinfachung der Berechnung dieses Schnittpunktes erforderlich. Aus Gleichung (5) wird ersichtlich, dass der Schnittpunkt x_{12} vom Verhältnis k_{sp}/k_{sn} (proportional zu w_p/w_n), den normierten Einsatzspannungen n und p sowie von dem Term $k_{lp} \tau_{in}/(C_L V_{dd}) =: a$ abhängt. Durch die Darstellung der numerischen Lösung (rote Kurve in Abb. 4) kann gezeigt werden, dass eine abschnittsweise Approximation (grüne Kurven) mittels einer Parabel und einer verschobenen $1/x$ -Funktion durch Anpassung dreier Parameter möglich ist.

Die für die Berechnung des weiteren Signalverlaufs notwendigen Lösungen der Differentialgleichung (3) können analytisch abgeleitet werden. Dies ist für die Regionen 2 und 4 jeweils eine Parabelfunktion, sowie für die Region 6 eine Geradengleichung.

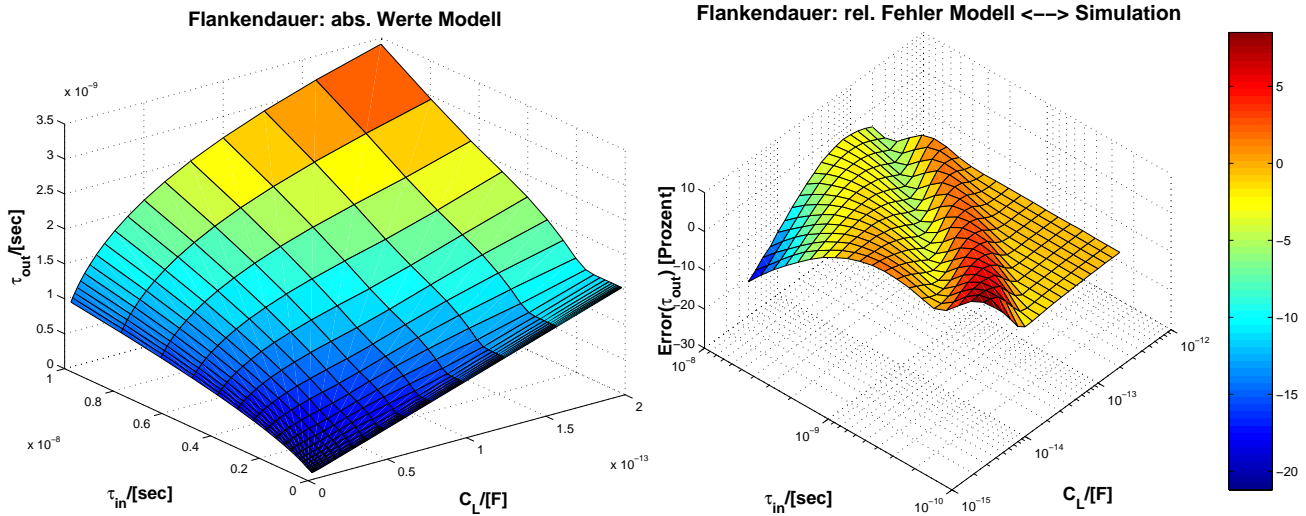


Abbildung 6. Flankendauer des Ausgangssignals: absolute Werte und relativer Fehler im Vergleich zur Simulation.

4 Ergebnisse

Mit obigen Ergebnissen ist es nun möglich die Verzögerungszeit, die Flankendauer des Ausgangssignals sowie die Verlustleistung eines Schaltvorgangs an einem Inverter ohne rechenintensive Operationen zu ermitteln. Abbildung 5 zeigt für die Verzögerungszeit den absoluten Fehler zwischen dem Modell und den aus einer Analogsimulation gewonnenen Referenzdaten.

Für schnelle und damit steile Eingangsflanken weist das abgeleitete Modell einen Fehler von ca. 10 ps auf, was in Bezug auf die minimale Anstiegszeit der vorliegenden Technologie ($\tau_{rise} = 90$ ps) sehr gering ausfällt. Für sehr langsame und somit flache Eingangsflanken steigt der Fehler des Modells an. In diesem quasi-statischen Fall haben die Abweichungen des MOS-Transistormodells einen nicht unwesentlichen Einfluss auf das Ergebnis. Der absolute Fehler mit 100 ps beträgt im Verhältnis zur Dauer der Eingangsflanke von bis zu 10 ns nur 1%.

Die aus dem Modell errechnete Flankendauer des Ausgangssignals und der relative Fehler zwischen Modell und Analogsimulation sind in Abb. 6 dargestellt. Bei den absoluten Werten der Flankendauer ist andeutungsweise ein Knick im Bereich kleiner Werte von τ_{in} erkennbar, der durch Tabellenmodelle problematisch zu erfassen ist. Die Darstellung des relativen Fehlers zeigt eine sehr gute Übereinstimmung zwischen Modell und Simulation mit einem Fehler <5%. Die Ausnahme bildet wiederum der quasi-statische Fall mit langsamer Eingangsflanke und kleiner Lastkapazität.

Für die in Abb. 7 dargestellte Verlustleistung zeigt sich eine ähnlich gute Übereinstimmung. Die Verlustleistung setzt sich hierbei sowohl aus einem zum Umladen der Lastkapazität notwendigen Anteil als auch aus dem Beitrag des Kurzschlussstroms zusammen.

Verlustleistung: rel. Fehler Modell <--> Simulation

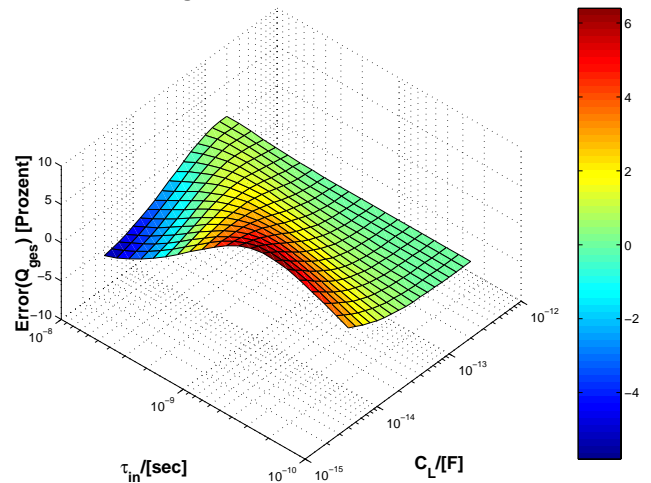


Abbildung 7. Relativer Fehler der Verlustleistung.

5 Zusammenfassung

Die Verzögerungszeit, die Ausgangsflankensteilheit und der Querstrom eines CMOS-Gatters werden sowohl durch die Lastkapazität als auch durch die Steilheit des Eingangssignals beeinflusst. Als Alternative zu den heute verwendeten Tabellenmodellen können analytische Modelle mit vergleichbarem Rechenaufwand aber ohne Interpolationsfehler diese Daten liefern. In dem hier gezeigten Modell ist es unter Beibehaltung des Verhältnisses w_p/w_n der Transistorweiten sogar möglich, beliebig dimensionierte Inverter zu analysieren. Insbesondere in Verbindung mit einer Generation von Treiberzellen "on-the-fly" könnten bei der Optimierung von Taktsystemen bessere Ergebnisse erzielt werden als bei einer Restriktion auf eine vordefinierte Menge unterschiedlich dimensionierter Treiberzellen. Hinsichtlich der Genauigkeit liegen die Ergebnisse des hier dargestellten Modells im Ver-

gleich zu Referenzwerten aus SPICE Simulationen bei typischerweise 5%.

Literatur

- Auvergne, D., Daga, J. M., and Rezzoug, M.: Signal Transition Effect on CMOS Delay Evaluation, *IEEE Trans. on Circuits and Systems – I: Fundamental Theory and Applications*, Vol. 47, No. 9, 1362–1369, 2000.
- Bisdounis, L., Nikolaidis, S., and Koufopavlou, O.: Analytical Transient Response and Propagation Delay Evaluation of the CMOS Inverter for Short-Channel Devices, *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 2, 302–306, 1998.
- Bisdounis, L. and Koufopavlou, O.: Short-Circuit Energy Dissipation Modeling for Submicrometer CMOS Gates, *IEEE Trans. on Circuits and Systems – I: Fundamental Theory and Applications*, Vol. 47, No. 9, 1350–1361, 2000.
- Dutta, S., Shetti, S. S. M., and Lusky, S. L.: A Comprehensive Delay Model for CMOS Inverters, *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 8, 864–871, 1995.
- Hedenstierna, N., and Jeppson, K. O.: CMOS Circuit and Buffer Optimization, *IEEE Trans. on CAD*, Vol. CAD-6, No. 2, 270–281, 1987.
- Hirata, A., Onodera, H., and Tamaru, K.: Estimation of Short-Circuit Power Dissipation for Static CMOS Gates, *IEICE Trans. Fund. Electron., Commun. Comput. Sci.*, vol. E79-A, No. 3, 304–311, 1996.
- Jeppson, K. O.: Modeling the Influence of the Transistor Gain Ratio and the Input-to-Output Coupling Capacitance on the CMOS Inverter Delay, *IEEE Journal of Solid-State Circuits*, Vol. 29, No. 6, 646–654, 1994.
- Sah, C. T.: Characteristics of the Metal-Oxide-Semiconductor Transistors, *IEEE Trans. on Electron Devices*, Vol. ED-11, 324–345, 1964.
- Sakurai, T. and Newton, A. R.: Alpha-Power Law MOSFET Model and its Application to CMOS Inverter Delay and Other Formulas, *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 2, 584–594, 1990.
- Sakurai, T.: Delay Analysis of Series-Connected MOSFET Circuits, *IEEE Journal of Solid-State Circuits*, Vol. SC-26, No. 2, 122–131, 1991.
- Velghe, R. M. D. A., Klaasen, D. B. M., and Klaasen, F. M.: Compact MOS modeling for analog circuit simulation, *IEDM Technical Digest*, 485–488, 1993.
- Vemuru, R. and Scheinberg, N.: Short-Circuit Power Dissipation Estimation for CMOS Logic Gates, *IEEE Trans. on Circuits and Systems – I: Fundamental Theory and Applications*, Vol. 41, No. 11, 762–765, 1994.